

Single-Image Shape from Defocus

José R.A. Torreão and João L. Fernandes
Instituto de Computação
Universidade Federal Fluminense
24210-240 Niterói RJ, BRAZIL

Abstract

The limited depth of field causes scene points at various distances from a camera to be imaged with different amounts of defocus. If images captured under different aperture settings are available, the defocus measure can be estimated and used for 3D scene reconstruction. Usually, defocusing is modeled by gaussian convolution over local image patches, but the estimation of a defocus measure based on that is hampered by the spurious high-frequencies introduced by windowing. Here we show that this can be ameliorated by the use of unnormalized gaussians, which allow defocus estimation from the zero-frequency Fourier component of the image patches, thus avoiding spurious high frequencies. As our main contribution, we also show that the modified shape from defocus approach can be extended to shape estimation from single shading inputs. This is done by simulating an aperture change, via gaussian convolution, in order to generate the second image required for defocus estimation. As proven here, the gaussian-blurred image carries an explicit depth-dependent blur component - which is missing from an ideal shading input -, and thus allows depth estimation as in the multi-image case.

1. Introduction

Shape from defocus (SFD) concerns the estimation of shape from the variable degree of blurring which results from the limited depth of field of imaging systems. This causes scene points at various distances from the camera to be imaged with different amounts of defocusing. If the defocus measure is estimated - which requires two or more images acquired under different aperture settings -, a depth map of the scene can be easily inferred from geometrical optics.

Usually, the defocusing process is modeled via the convolution of the perfectly focused image with a point spread function (PSF), whose spatial dimension is proportional to the defocus parameter. This is not a strictly realis-

tic approach, since defocusing is a spatially varying process. Nevertheless, assuming a gaussian PSF, Pentland introduced a Fourier-domain algorithm for defocus estimation where convolution over local patches is assumed, in order to account for depth-dependent blur [1]. Other frequency- and spatial-domain techniques followed, most of them sharing the local convolution assumption [2]-[5], but the problems entailed by the use of local windows were soon identified, such as the introduction of spurious high-frequency components - due to the artificial discontinuities at window boundaries -, and irradiance leaking through neighboring patches.

Here we modify Pentland's formulation of the convolution-based SFD, by modeling the defocusing process through the convolution with unnormalized, instead of normalized, gaussians. Besides seeming more appropriate to deal with irradiance bleeding across neighboring windows, this model allows us to estimate the defocus measure from the zero-frequency (DC) Fourier component of the image patches, thus avoiding the aforementioned spurious high-frequency components. Once the defocus measure has been thus estimated, the 3D reconstruction of the imaged surfaces proceeds as usual, from geometrical optics.

As our main contribution, we also show that the modified SFD approach can be extended to shape estimation from single shading inputs. Ideal shading images, as used for shape-from-shading estimation, carry no depth-dependent blur, since they are assumed as having been captured under orthographic projection. Here we model the formation of such a uniformly blurred image, by introducing an overall defocus measure consisting of the product of a depth-dependent blur component and an (spatially-varying) aperture blur component which compensates the former. We then proceed to show that, by simulating an aperture change through the convolution of the shading image with a gaussian PSF, the depth-dependent blur component can be made explicit in the new image, and estimated via the modified SFD approach. Shape reconstruction may then proceed as in the multi-image case.

The remainder of this article is organized as follows: In

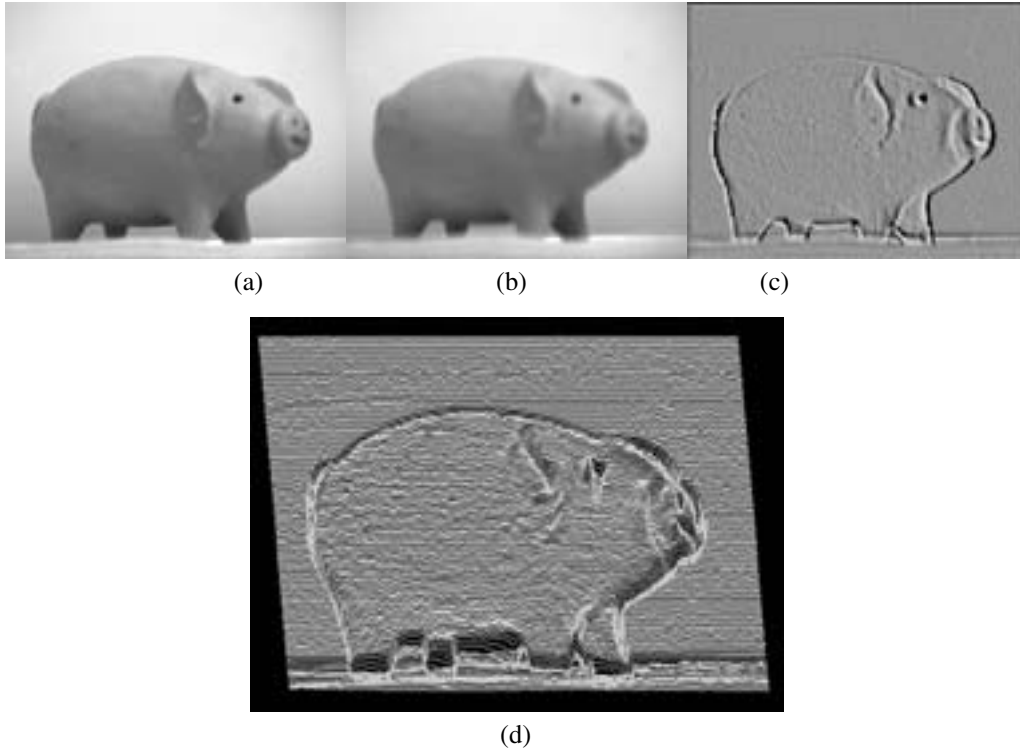


Figure 1. (a) and (b) Pig images captured under two different aperture settings; (c): Rendition of the estimated surface function in (d), for lambertian reflectance and uniform albedo, with illumination from (1,1,1).

Section 2, we introduce our model for defocus estimation based on local convolutions by unnormalized gaussians, and show an example of its application. The process is then extended to single-image shape estimation in Section 3, and illustrated by experimental results. The article then concludes with our final remarks in Section 4.

2. Shape from Defocus

As shown in [1], the relation between the depth map and the defocus map of a scene can be derived from geometrical optics, as

$$Z(x, y) = \frac{Fv_0}{v_0 - F - fR(x, y)} \quad (1)$$

where $R(x, y)$ denotes the local blur radius, due to defocusing, and where F , f and v_0 are parameters of the imaging system: respectively, the focal distance, the f-number (defined as the ratio between the focal distance and the aperture radius, $f = F/r$), and the image plane position relative to the focusing lens. According to (1), shape can be estimated from the defocusing measure R , which, under the gaussian

blur model, is identified as the spatially-varying standard deviation parameter,

$$R(x, y) = \sigma(x, y) \quad (2)$$

Also in [1], an approach for the estimation of $\sigma(x, y)$ has been proposed, based on the ratio of the Fourier transforms of corresponding patches in two images captured under different focus settings. Here we will take up essentially the same approach, with an important distinction: differently from [1], the defocusing of the local image patches will here be modeled through unnormalized gaussian filters. Namely, we assume that the intensities in a local image window can be expressed as the convolution

$$I(x, y) = \exp \left[-\frac{(x^2 + y^2)}{2\sigma^2} \right] * L(x, y) \quad (3)$$

where $L(x, y)$ is the ideally focused image, and σ denotes the local defocusing measure, assumed uniform over each window. Since there will always be some irradiance bleeding, due to the defocusing blur, across the borders of neighboring patches, it is reasonable not to assume normalization in this case.

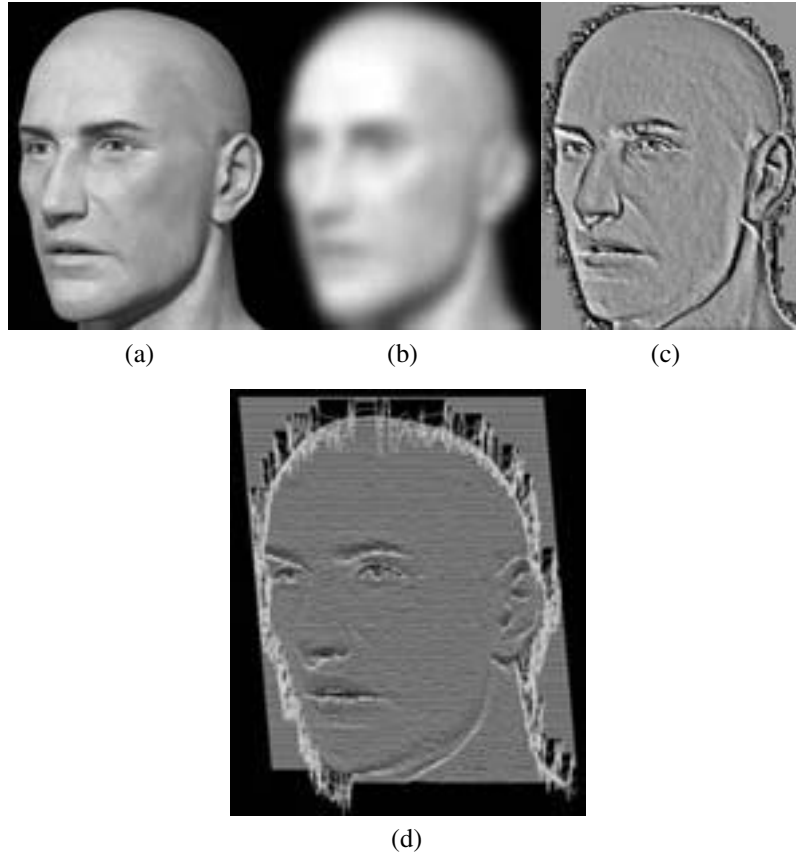


Figure 2. (a) Synthetic face image; (b) Gaussian blurred version of (a), for $\Delta\sigma = 0.05$; (c): Rendition of the estimated surface function in (d), for lambertian reflectance and uniform albedo, with illumination from (1,1,1).

Now, in the frequency domain, equation (3) becomes

$$\tilde{I}(\omega) = 2\pi\sigma^2 \exp\left[-\frac{\sigma^2\omega^2}{2}\right] \tilde{L}(\omega) \quad (4)$$

where we are using the tilde to denote the Fourier transform of the spatial signals, and where $\omega \equiv \sqrt{\omega_x^2 + \omega_y^2}$ is the spatial frequency magnitude. Therefore, if we consider corresponding local patches over a pair of images, say I and I_0 , captured under different focus settings, we obtain from (4)

$$\frac{\tilde{I}(\omega)}{\tilde{I}_0(\omega)} = \frac{\sigma^2}{\sigma_0^2} \exp\left[-\frac{\omega^2}{2}(\sigma^2 - \sigma_0^2)\right] \quad (5)$$

and thus,

$$\log \frac{\tilde{I}(\omega)}{\tilde{I}_0(\omega)} = \log \frac{\sigma^2}{\sigma_0^2} - \frac{\omega^2}{2}(\sigma^2 - \sigma_0^2) \quad (6)$$

Since such relation is assumed to hold at all frequencies, for the same local defocus measures, σ and σ_0 , we may

consider it at $\omega = 0$, to obtain

$$\sigma = \sigma_0 \sqrt{\frac{\tilde{I}(0)}{\tilde{I}_0(0)}} \quad (7)$$

The above, when computed over a series of local windows spanning the whole image will define the position-dependent defocus measure of equation (2). This, in turn, can be used for shape estimation via equation (1), which can be easily recast in a form that expresses the surface function, up to a multiplicative factor, in terms of the defocus measure and a single free parameter, here chosen empirically. The process is illustrated by the experiment in Fig. 1, where 2×2 windows have been used for defocus estimation.

3. Single-Image Approach to SFD

Given a single *shading* image, we now propose shape estimation based on a simulation of defocusing.

Ideal shading images, as used for shape from shading (SFS), are assumed to have been captured under the general reflectance-map conditions, including orthographic projection, and so they carry no depth-related blur. Shape from defocus can not therefore be based on them. On the other hand, since shading is due to surface orientation, and thus depth variation, *real* shading images, captured under finite aperture, must necessarily present spatially-variant, depth-dependent defocus, and so are amenable to SFD estimation; in such case, reflectance-map based shape from shading should work only approximately. This intrinsic contradiction between SFS and SFD can be accommodated if we assume that the uniformly defocused image, as required by SFS, has been captured under a finite position-dependent aperture condition, in such a way that the aperture- and depth-related blur effects compensate each other.

Let us illustrate this by considering a simple model, whereby we express the overall defocus measure as the product

$$\sigma = \sigma^a \sigma^d \quad (8)$$

where both σ^a and σ^d - respectively, the aperture- and depth-dependent blur components - could be functions of position. Such multiplicative model is consistent with geometrical optics, from whence we learn that the defocusing blur radius is proportional to the product of the aperture radius and the displacement of the imaged point from its perfectly focused position (equation (1) is ultimately a consequence of this [7]).

What we propose, then, is to model the input image, I_0 , as having been captured under a position-dependent aperture condition, such that

$$\sigma^a(x, y) = \frac{\sigma_0}{\sigma^d(x, y)} \quad (9)$$

for a constant σ_0 , such that I_0 appears under overall uniform blur.

Now, the effect of a uniform change in aperture (*irising*) of the imaging system can be simulated by convolving I_0 with a gaussian kernel, thus generating a new image, I . According to (8), if $\Delta\sigma$ is the additional blur entailed by the aperture change, the defocus measure of I will be given as

$$\sigma(x, y) = [\sigma^a(x, y) + \Delta\sigma]\sigma^d(x, y) = \sigma_0 + \Delta\sigma\sigma^d(x, y) \quad (10)$$

Equation (10) means that the depth-dependent blur component, σ^d , implicit in I_0 , becomes explicit in I , the artificially defocused image. In particular, if $\Delta\sigma \gg \sigma^a(x, y)$, the defocusing of I will be proportional to the depth-related blur.

From the Fourier transforms of I and I_0 over corresponding local windows, the position-dependent measure $\sigma(x, y)$ can be estimated through (7), thus allowing scene recon-

struction via equation (1), which now takes the form

$$Z(x, y) = \frac{Fv_0}{v_0 - F - F(\sigma_0/\Delta\sigma)\sqrt{\tilde{I}(0)/\tilde{I}_0(0)}} \quad (11)$$

where we have used $f = F/r$, and taken the aperture radius as equal to the aperture-component blur ($r = \sigma^a + \Delta\sigma$), also assuming $\Delta\sigma \gg \sigma^a(x, y)$. In (11), it is understood that $\tilde{I}(0)$ and $\tilde{I}_0(0)$ are window estimates, and thus functions of x and y .

It is worthwhile remarking that the expression for $Z(x, y)$ obtained above has a similar functional form as that which results from an alternative approach to SFS, that of the Green's function photometric motion (GPM) [6]. GPM can be related to the single-image SFD, inasmuch as both processes share the basic approach of generating an artificial pair to the shading input. While here we simulate an aperture change of the imaging system via convolution with a gaussian kernel, GPM employs the Green's function of a matching equation to simulate motion. In such case, the depth map of the imaged surface can be estimated as

$$Z(x, y) = \frac{k_1(1 + \gamma^2)u}{u'[k_0 - I(x, y)]} \quad (12)$$

where k_0 and k_1 are parameters of the linearized irradiance function, $I(x, y) = k_0 + k_1(p + \gamma q)$, and where $\gamma = \tan \theta$, for $\pi/2 - \theta$, denotes the direction of the simulated rotation. The remaining parameters, u and u' , are also constants, representing, respectively, the x -component of the optical flow and its directional derivative, $(\partial_x + \gamma\partial_y)u$, at $x = 0$.

All the other terms being constant, if (11) and (12) are to represent the same depth map, we should have

$$\sqrt{\tilde{I}(0)/\tilde{I}_0(0)} \propto I(x, y) = k_0 + k_1(p + \gamma q) \quad (13)$$

and thus

$$\tilde{I}(0) \propto \tilde{I}_0(0)[k_0^2 + 2k_0k_1(p + \gamma q) + k_1^2(p + \gamma q)^2] \quad (14)$$

Equation (14) can be interpreted as a quadratic image irradiance equation for $\tilde{I}(0)$ - which represents essentially the mean value of the intensities of I over each local window -, with $\tilde{I}_0(0)$ as a modulation factor. It should be recalled that the input image $I_0(x, y)$ is assumed to be an ideal SFS image, carrying only uniform defocus. The blurred image, $I(x, y)$, on the other hand, has its depth-dependence made explicit by the term under the square brackets in (14).

It is also worth remarking that an alternative interpretation can be given to I_0 - if we assume it as having been captured under *fixed* and finite aperture -, as the representation of a flat scene, and thus of an albedo function. Equation (14) then displays the standard form for an image irradiance equation, consisting of the product of a position-dependent albedo and a reflectance-map function.

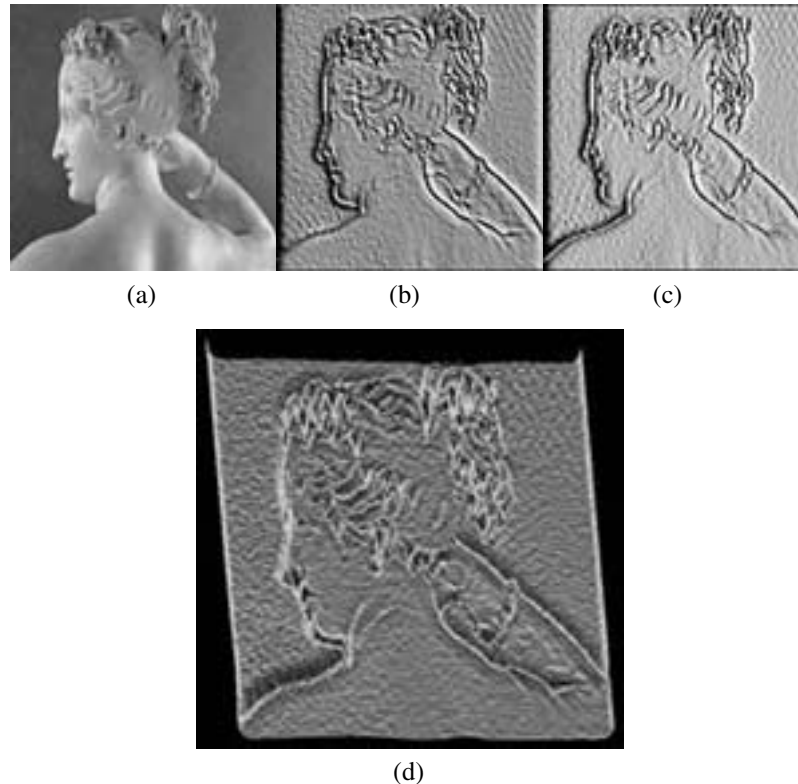


Figure 3. (a) Paolina image; (b) and (c): Renditions of the estimated surface function in (d), for lambertian reflectance and uniform albedo, with illumination from (1,1,1) and (1,0,1).

Examples of the application of the single-image SFD approach appear in Figs. 2 to 4. In all cases, given the input image, a uniformly blurred version of it was generated via gaussian convolution (with $\Delta\sigma = 0.05$), as illustrated by Fig. 2b. The defocus measure was then estimated in 2×2 windows through equation (7), and used, for reconstruction purpose, in equation (11). As in the two-image experiment of Fig. 1, the depth function was estimated up to a multiplicative factor, and with its single free parameter empirically chosen.

4. Concluding Remarks

The following contributions have been reported here, to the process of shape estimation from defocus (SFD):

i) We have modified Pentland's defocus estimation approach [1], which is based on gaussian-convolution over local image patches, by considering unnormalized, instead of normalized gaussians. Besides being more appropriate to account for irradiance bleeding, the use of unnormalized gaussians allows the estimation of defocus from the DC Fourier component of the image patches, thus avoiding

the undesirable high-frequency components introduced by windowing.

ii) We have presented a single-image shape from defocus process, which is based on simulating an aperture change of the imaging system, in order to generate the second image required by the defocus estimation approach. We introduced a multiplicative model for the defocus measure - expressing it as the product of an aperture-dependent and a depth-dependent factor -, and assumed that the shading input image, which carries no depth-dependent blur, has been acquired in such a way that those two factors compensate each other. Based on such model, we have then been able to show that the gaussian-blurred image carries the depth-dependent defocus information which is missing from the input. Such information can be estimated via our modified shape-from-defocus approach, and surface reconstruction then proceeds as in the multi-image case.

The single-image SFD introduced here follows the same line of approach which led to our previous Green's function shape from shading (GSFS) [8, 9] and Green's function photometric motion (GPM) [6], where an artificial pair to the single input is generated by simulating a certain photometric or geometric intervention on the imaging set-up. In GSFS and GPM, Green's functions of image matching

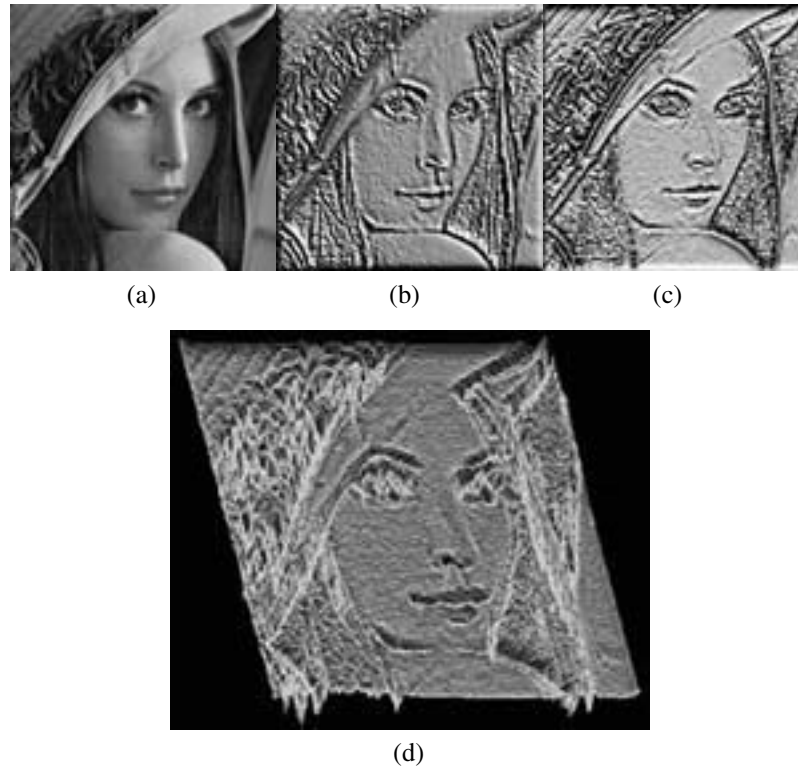


Figure 4. (a) Lenna image; (b) and (c): Renditions of the estimated surface function in (d), for lambertian reflectance and uniform albedo, with illumination from $(-1,-1,1)$ and $(0,1,1)$.

equations are used for simulating, respectively, a change of illumination and a rotation of the imaged surface; in the single-image SFD, on the other hand, a change in aperture is simulated via convolution with a gaussian kernel. A remarkable finding of the present work is that the depth map obtained through the latter approach shows the same functional form as that yielded by GPM. Not less because they are essentially distinct and unrelated processes, we believe that such result adds to the credibility of both.

References

- [1] Pentland, A.P. (1987). A new sense for depth of field, IEEE Trans. PAMI 9(4): 523-531.
- [2] Xiong, Y. and Shafer, S.A. (1993). Depth from focusing and defocusing, in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition: 68-73.
- [3] Ens, J. and Lawrence, P. (1993). An investigation of methods for determining depth from focus, IEEE Trans. PAMI 15(2): 97-108.
- [4] Pentland, A.P., Scherok, S., Darrell, T., and Girod, B. (1994). Single range camera based on focal error, J. Opt. Soc. Am. A11: 2925-2934.
- [5] Ziou, D. and Deschenes, F. (2001). Depth from defocus in spatial domain, Comp. Vision and Image Understanding 81: 143-165
- [6] Torreão, J.R.A. and Fernandes, J.L. (2004). From photometric motion to shape from shading, in Proceedings of SIBGRAPI 2004, IEEE Computer Society, Los Alamitos, USA: 186-191.
- [7] Horn, B.K.P. (1986). *Robot Vision*. Cambridge, MA: MIT Press
- [8] Torreão, J.R.A. (2001). A Green's function shape from shading, Patt. Recognition 34: 2367-2382.
- [9] Torreão, J.R.A. (2003). Geometric-photometric approach to monocular shape estimation, Image and Vision Computing 21: 1045-1061.